### The Problem

The concept of FAIR digital object is quite wide and can vary in term of granularity and type of data that should be represented.

In fact, a FAIR digital object can be a single atomic information (e.g., a coded diagnosis) up to a collection of data (e.g., a data set). Data, moreover, can represent quite different kinds of information: it might be for example a waveform, an image, a condition, a medication, or other kinds of data.

This variety makes not so straightforward the mapping between FAIR data objects and the FHIR resources covering a large range of resources and elements (see figure below).

|  |
| --- |
|  |

Similar considerations can be done also for the metadata, where, depending on what the metadata is, the metadata elements can be represented within the same resource documenting the data, or by a set of linked resources.

|  |
| --- |
|  |

#### Data vs metadata

In addition to the issue reported above, depending on what a "FAIR object" is the boundary between data and metadata may change. That means that based on the context, the same information can be considered as part of the data or of the metadata. This makes even more complex the mapping to the FHIR resources.

The figure below shows how information that were part of the metadata in the previous example are now part of the data; and how metadata and data elements are represented in the same FHIR resource.

|  |
| --- |
|  |

#### Metadata identification

In consideration of the previous arguments, the expectation of having a persistent and unique ID that identifies the metadata FAIR data object, distinct from the data FAIR data object, cannot be always satisfied in the FHIR space, except for specific contexts.

### The approach

Even though many levels of granularity can be considered in this work we distinguish two main levels:

1. The **study level**: describing the collection of data referring to a specific study, publication, usage context.
2. **Instance** (or patient) **Level**: providing the metadata/data information associated to a single subject.

|  |
| --- |
|  |
| **Figure 1 - Level of data objects granularity considered in this guide.** |

If the boundary between metadata and data is quite clear for the collection (study level)

|  |
| --- |
|  |
| **Figure 2 – Metadata and data for a data collection.** |

This might be not so true for the instance level, where information as the gestational week or the mother height and weight, might be considered metadata for the signals, but also patient level data.

|  |
| --- |
|  |
| **Figure 3 – Metadata and data for the instance level data.** |

### Metadata Representation in FHIR

#### Study Level (Collection)

HL7 FHIR provides several candidates resources that could be used to represent collection of data (e.g., Bundle, Lists..); some of them however seem not to fully support the “rich metadata” FAIR requirement.

At this stage three FHIR resources have been identified to represent Collection metadata:

* Library
* Citation
* ResearchStudy

The choice of the resource depends on the usage context.

|  |
| --- |
|  |
| **Figure 4 – Data Collection Metadata representation** |

The figure highlight the version of FHIR from which this resource is available and how the linkage between metadata and data is realized (with the FHIR version where this is derived from).

Considering three kinds of possible data objects:

1. Non-FHIR object
2. FHIR and non-FHIR objects
3. FHIR-only objects

The following table highlight how this linkage could be realized.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Library content** | **Citation relatedTo.target[x]** | **Research Study results** |
| **Applicable to** | FHIR R4 | FHIR R5 (R4b) | FHIR R5 |
| **Non-FHIR binary** | link via uri | link via uri | EvidenceReport.relatedArtifact |
| **Combination of FHIR and non-FHIR objects** | link via uri  *(Add an extension ?)* | Reference to a FHIR resource | EvidenceReport.relatedArtifact |
| **Only FHIR** | link via uri  *(Add an extension ?)* | Reference to a FHIR resource | EvidenceReport.relatedArtifact |

*< to be continued>*

### Conclusions

*< move to another page ?>*

* Do not pretend to address all the identified issues and realize all the FAIR recommendations since the beginning: **FAIRness is a continuous improvement process**.
* Follow a progressive approach focusing on the characteristics that are relevant for your context, balancing advantages and costs. For example, you may reasonably accept that metadata and data information at the patient level may be represented by a set of already existing FHIR resources, focusing your effort on a well identifiable and distinct representation of the metadata of the data collection at the study level.